Identifying the RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RDBmap

Alfredo Castello[1, 2,*], Christian K. Frese[1,3], Bernd Fischer[1, 4, 6], Aino I. Järvelin[2], Rastislav Horos[1], Anne-Marie Alleaume[1], Sophia Foehr[1], Tomaz Curk[1, 5], Jeroen Krijgsveld[1, 4], Matthias W. Hentze[1,*]

1. European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany
2. Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK
3. CECAD Research Center, University of Cologne, Joseph-Stelzmann-Str. 26, 50931 Cologne, Germany
4. German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
5. Faculty of Computer and Information Science, University of Ljubljana, 1001 Ljubljana, Slovenia
6. Deceased

*Correspondence to:

hentze@embl.de and alfredo.castellopalomares@bioch.ox.ac.uk

KEYWORDS: RNA-binding protein; RNA-binding domain; protein-RNA interaction; RNA; RNA interactome; UV crosslinking.

KEY REFERENCES:

Comprehensive Identification of RNA-Binding Domains in Human Cells. Castello A, Fischer B, Frese CK, Horos R, Alleaume AM, Foehr S, Curk T, Krijgsveld J, Hentze MW. Mol Cell. 2016 Aug 18;63(4):696-710. doi: 10.1016/j.molcel.2016.06.029. PMID: 27453046

The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. Liao Y, Castello A, Fischer B, Leicht S, Föehr S, Frese CK, Ragan C, Kurscheid S, Pagler E, Yang H, Krijgsveld J, Hentze MW, Preiss T. Cell Rep. 2016 Aug 2;16(5):1456-69. doi: 10.1016/j.celrep.2016.06.084. PMID: 27452465

EDITORIAL SUMMARY: Here the Authors provide an extension to their earlier RNA interactome capture protocol, describing RBDmap; a method to identify the regions of RNA-binding proteins (RBPs) engaged in native interactions with RNA, in a proteome-wide manner.

TWEET: RBDmap - an extension of the RNA Interactome Capture method to identify RNA-binding domains of RBPs in a proteome-wide manner

@Alf_Castello

SUMMARY

RBDmap identifies, in a proteome-wide manner, the regions of RNA-binding proteins (RBPs) engaged in native interactions with RNA. In brief, cells are irradiated with UV light to induce protein-RNA crosslinks. Resulting covalently linked protein-RNA complexes are purified with oligo(dT) magnetic beads, following stringent denaturing washes. After elution, RBPs are subjected to partial proteolysis, where the protein regions still bound to the RNA and those released to the supernatant are separated by a second oligo(dT) selection. After sample preparation and mass spectrometric analysis, peptide intensity ratios between the RNA-bound and released fractions are

used to determine the RNA-binding regions. As a Protocol Extension, this article describes an adaptation of an existing Protocol, and offers additional applications. The earlier protocol (for the RNA interactome Capture method) describes how to identify the active RBPs of cultured cells, whilst this Protocol Extension enables the identification of the RNA-binding domains of RBPs in cultured cells. The experimental workflow takes one week, plus two additional weeks for proteomics and data analysis. Notably, RBDmap presents numerous advantages over classical methods to determine RNA-binding domains: it produces proteome-wide, high resolution maps of the protein regions contacting the RNA in a physiological context and can be adapted to different biological systems and conditions. Because RBDmap relies on the isolation of polyadenylated RNA via oligo(dT), it will not provide RNA-binding information on proteins interacting exclusively with non-polyadenylated transcripts. Applied to HeLa cells, RBDmap uncovered 1,174 RNA-binding sites in 529 proteins, many of which were previously unknown.

INTRODUCTION

RNA is a key molecule in cell biology that displays a broad variety of functions, including not only its well-studied role as a template of protein synthesis (mRNA) but also its activity as structural component of cellular machines (e.g. rRNA), regulator of gene expression (e.g. miRNA and non-coding RNA) and catalytic core (e.g. ribozymes). RNA-binding proteins (RBPs) associate with RNA from synthesis to decay forming dynamic ribonucleoproteins (RNPs) that regulate RNA fate. In spite of the relevance of RBPs in the RNA's life, the diversity of cellular proteins enabled with RNA-binding activity remained largely unknown. The most well-characterized RBPs,

referred to here as ‘classical’, are distinguished by the presence of one or multiple well-known RNA-binding domains (RBDs), such as the RNA recognition motif (RRM), the K-homology domain (KH), cold shock domain (CSD), the DEAD box helicase domain, the double stranded RNA-binding domain (dsRBD) and others (see complete list in[1]). Interestingly, the architecture of classical RBPs often follows a modular design combining multiple, typically globular, RBDs within the same protein to increase RNA-binding selectivity and affinity[1,2].

In the last three decades, numerous RBPs lacking classical RBDs were discovered in a stepwise fashion, raising the possibility of unexplored modes of RNA-binding[3-7]. Recent work from our laboratory and others combined *in vivo* ultraviolet (UV) crosslinking, stringent oligo(dT) isolation of polyadenylated RNA and quantitative mass spectrometry to determine the repertoire of cellular RBPs[8,9]. This method, called RNA interactome capture[10], added hundreds of non-canonical RBPs to the previously known repertoire of RBPs, including metabolic enzymes, protein chaperons, and cytoskeleton-associated proteins[8,9]. RNA interactome capture has recently been applied to other cell lines and organisms, including *Saccharomyces cerevisiae*, *Plasmodium falciparum*, *Caenorhabditis elegans*, *Drosophila melanogaster* and plants, providing unprecedented insights into the diversity of RBPs and their evolution across species[11-23]. RNA interactome capture has also been combined with subcellular fractionation, implementing a DNase step to reduce the incidence of DNA in the nuclear preparation[21]. Many of the RBPs discovered by these studies have been shown to play important roles in RNA biology[13,24-27] and many others are currently under investigation. However, it remained largely unknown how these novel RBPs interact with RNA.

Classical methods to identify RBDs effectively determined the regions involved in RNA binding using individual proteins and deletion or substitution mutagenesis (e.g.[28]). However, these methods are unsuitable for proteome-wide RBD discovery, calling for different experimental approaches. Here we describe RBDmap, an extension of the RNA interactome method[10], that combines UV crosslinking, oligo(dT) capture, controlled proteolysis and mass spectrometry to discover high-confidence RNA-binding sites on a proteome-wide scale (Figure 1). Applied to HeLa cells, RBDmap identified 1174 binding sites within 529 human RBPs, including both classical and unorthodox RBPs[29]. RBDmap has also proven its utility in HL-1 cardiomyocytes, where RBDmap dissected the RNA-binding regions of 368 murine proteins[30].

Overview of the protocol

RBDmap (Figure 1) employs two alternative UV crosslinking approaches to immobilize native protein-RNA interactions in cultured cells. The first approach, referred to as conventional UV crosslinking (cCL), exploits the intrinsic reactivity of nucleotide bases to irradiation with UV light at 254 nm. At this wavelength, UV irradiation induces short-lived, free radicals at the nucleotide bases that attack amino acids at “zero” distance to form covalent bonds[31]. In parallel, we used photoactivatable ribonucleoside–enhanced crosslinking (PAR-CL) that employs 4-thiouridine (4SU). This nucleotide analogue is taken up by cultured cells and incorporated into nascent RNAs. Crosslinking is achieved after irradiation with 365 nm UV light[8,9]. Since the absorbance of natural bases at this wavelength is negligible, protein-RNA crosslinks are 4SU-dependent. Since RBPs can differ in their crosslinking efficiencies with either of the two protocols[9], we applied both cCL and PAR-CL in parallel to enhance RBD

identification. After denaturing cell lysis, polyadenylated RNA and its covalently bound proteins are purified under stringent washing conditions by hybridization with oligo(dT) coupled to magnetic beads. After elution at 55°C, RBP-RNA complexes are treated with a protease (LysC or ArgC) that cleaves every 17 amino acids on average. Proteolytic treatment generates two types of RBP fragments: i) peptides that remain crosslinked to the RNA and include the RNA-binding site; and ii) peptides that are released into the supernatant which correspond to protein regions that are not bound (or crosslinked) to RNA, referred to as “released peptides” (Figure 1). Both classes of protein fragments are separated by a second oligo(dT) isolation, where the supernatant contains the released protein fragments and the bead-retained fraction is enriched for RNA-bound fragments (Figure 1). Both the supernatant and the elution of the second oligo(dT) capture are treated with RNases and trypsin, and subsequently analyzed by mass spectrometry (MS). Because trypsin cleaves every 8 amino acids (aa) on average, the ~17 aa RNA-bound protein fragments will be cleaved into a peptide(s) that remains covalently linked to one or more nucleotides (X-link peptide) and its neighboring tryptic peptides (N-link) that are released (Figure 1). While the X-link peptide cannot be readily identified due to the nucleotide remnant[32,33], the N-link peptide(s) can be identified by conventional MS. RNA-binding sites are determined by comparing the ratio of each N-link peptide in the RNA-bound versus the released fraction, applying a 1% false discovery rate (FDR) as the statistical cut off (Figure 2A). Peptides classified as “RNA-binding” correspond to the MS-identifiable N-link moiety, and their boundaries are extended in silico to the nearest LysC or ArgC cleavage sites to recapitulate the original fragment generated by these proteases, which contains the actual RNA-binding site. These fragments with an average length of 17 aa are referred to as “RBDpeps”.

Advantages of the method in comparison to other approaches

RBDs are classically mapped by loss-of-function approaches, which involve the generation of single point or deletion mutants that are tested for RNA binding *in vitro* by EMSA assays[28] or *in vivo* by crosslinking and immunoprecipitation (CLIP) followed by T4 polynucleotide kinase (PNK) assay or fluorescence-based methods[11,34]. In contrast to CLIP followed by PNK labeling and classical EMSA assays, RBDmap does not require the use of radioactivity. While commonly successful, the limited number of mutations that can be generated and tested often impacts on the resolution of the RBD mapping, especially when the boundaries of the structural fold associated with RNA binding are not well defined. Mutations can cause protein misfolding, mislocalisation or dysfunction, limiting the utility of this approach. Furthermore, the work required for cloning, expression and purification limit this approach to a small number of cases at a time. RBDmap displays notable advantages over these classical approaches: i) UV crosslinking is applied to living cells, thus mapping native protein-RNA interactions; ii) protein overexpression is not required, saving time and avoiding artifacts associated with overexpression; iii) it provides information on hundreds of RBPs at a time; iv) it can be applied to different experimental conditions and biological systems; and v) it is unbiased regarding the identity and boundaries of the RBD.

More recently, Kramer *et al*. developed a sophisticated algorithm to identify peptides crosslinked to nucleotides by MS[33]. Applied to RNA interactome samples in *Saccharomyces cerevisiae*, this method uncovered 257 RNA-bound peptides in 124 RBPs. While this approach provided bona fide RNA-binding sites with very high resolution, it suffers from a bias towards domains that crosslink particularly well, such as RRMs. Conversely, RBDmap reports RNA-binding sites at ~17 amino acids resolution by identification of unmodified neighboring peptides that are efficiently

detected by mass spectrometry, even when present in limited amounts. RBDmap identified 1174 RNA-interacting regions from 529 Hela cell RBPs, mapping equally to classical and non-canonical RBDs[29]. He *et al.* recently described an alternative method called RBR-ID[35], which exploits the mass shift of peptides crosslinked to a nucleotide remnant in an indirect way. In brief, RBR-ID classifies peptide peaks with reproducibly lower intensity in UV irradiated samples than in non-irradiated controls as putative RNA-binding sites, assigning the loss of signal to the presence of crosslinked nucleotides that increase the peptide mass. RBR-ID complements RBDmap because it does not involve oligo(dT) selection and can thus assign RBDs to RBPs lacking association with polyadenylated RNAs. As a disadvantage, RBR-ID lacks a purification step to enrich for RBPs, which results in a highly complex mixture of peptides corresponding to the whole cell proteome. High variations in intensity of peptides with low abundance is a common phenomenon in highly complex samples and may yield a higher incidence of falsely positive results. As shown by in silico and experimental validation[29], the two purification steps of RBDmap ensure that i) most of the peptides identified correspond to RBPs and ii) those enriched in the RNA-bound fraction actually map to RNA-binding sites.

Limitations of the approach

RBDmap will not be suitable to identify RBDs in the following scenarios: i) the RBP of interest is not active or not expressed under the experimental conditions employed; ii) the geometry of the protein-RNA interaction is not suitable for UV crosslinking, typically including RBDs that bind to double-stranded RNA; iii) the RBP binds exclusively to non-polyadenylated RNAs; iv) the LysC or ArgC (RBDpep) fragments crosslinked to

the RNA lack internal cleavage sites for trypsin, leading to a lack of MS-identifiable N-link peptides; v) the available material is too limited (e.g. non-dividing primary cells).

These limitations could be circumvented as follows. Prior to using RBDmap, we recommend employing available RNA and protein expression atlases to identify the cell type/line or the physiological conditions where the protein of interest is expressed. Alternatively, the protein can be expressed using transient transfection or stable cell lines to maximize the chances of obtaining an RBDmap profile. Overexpression may facilitate non-physiological interactions. However, RNA-binding profiles obtained under these conditions may still be informative and the identified RNA-binding sites can subsequently be validated in a more physiological context. Inefficient crosslinking of the protein of interest can be overcome by small scale set up experiments using RNA interactome capture followed by western blotting[10]. Different UV-crosslinking methods (PAR-CL or cCL), UV wavelengths (254, 312 or 360 nm[35]) and UV dosages can be tested to tailor the protocol to a given RBP. If the RBP does not bind to poly(A) RNA, RBDmap could potentially be combined with sequence-specific probes instead of oligo(dT) to capture its target RNA (if known)[36-38]. As discussed below, RBPs with suboptimal sequence for LysC and ArgC cleavage can be digested with alternative proteases (Figure 2B). We recommend analysis of the protein sequence and cross-comparison with the known protease cleavage specificities prior to initiating an RBDmap experiment. Finally, limited starting material represents a challenge for RBDmap; however, new and more sensitive mass spectrometers are released every year. Sensitivity improvements will positively impact on the input requirements of RBDmap experiments.

Applications and future uses of the method

RBDmap was developed as a method for the system-wide identification of RBDs in living cells. The current repertoires of RBDs identified from HeLa[29] and HL-1 cells[30] can be expanded using other cell types and organisms to generate a catalogue of RBDs used by living systems. The successful implementation of RNA interactome capture in different cell lines (HeLa[9], Huh-7[13], HEK293[8], HL-1[30], RAW 264[39], mouse embryonic stem cells[11]) and organisms (plants[17-19], *Drosophila melanogaster*[15,16], *Caenorhabditis elegans*[14], *Saccharomyces cerevisiae*[12-14] and *Plasmodium falciparum*[20]) expands the scope of biological systems where RBDmap can be readily employed. It can thus be used to study the evolution of RBDs across species to provide new insights into the evolution and regulation of eukaryotic RBDs.

Moreover, RBDmap can further be adapted to focus on single RBPs by overexpressing the protein of interest in cultured cells to maximize protein coverage in RBDmap assays. Single protein applications can substitute the first oligo(dT) capture step by immunoprecipitation using antibodies against the protein of interest. This approach is complementary to conventional methods to map RBDs (see above), with the advantage of obtaining high resolution RNA-binding sites (~17 amino acids) without the need to generate a large repertoire of deletion or single point mutants.

Notably, RBDmap can be applied to different physiological conditions to determine, in a global and comparative manner, whether the RNA-binding properties of RBDs respond to biological cues. Such studies could shed light on regulatory mechanisms controlling RNA binding and their links to altered cellular states (e.g. stress, differentiation, activation, division or apoptosis). Because RNA-binding sites heavily overlap with phosphorylation, acetylation and methylation sites[29], it will be of great

interest to carry out such studies in parallel with proteome-wide analyses of posttranslational modifications performed under the same experimental conditions. These comparative analyses will shed light on the influence of posttranslational modifications in the regulation of RNA-binding**.**

Experimental design

Controls: the “RNA-bound” fraction eluted after the second round of oligo(dT) capture contains the protein fragments engaged in RNA binding (Figure 1). The supernatant of the second oligo(dT) capture contains the protein regions released from the RNA after the protease treatment, referred to here as “released” fraction (Figure 1). These two fractions are critical for the determination of RNA-binding sites, which are calculated using the peptide intensity ratios between the RNA-bound and the released fractions. RBDpeps are called by computational extension of the N-link peptide to the adjacent LysC or ArgC cleavage sites. The released fraction is thus a suitable negative control against which to test the intensities of every N-link peptide identified in the RNA-bound fraction. Because tryptic digestion is required to release the N-link peptide from the Lys/ArgC RBDpep, omitting trypsin will render no identifications in the RNA-bound fraction. This potential control can be included at Step 33. Because of the lack of trypsin-derived neighboring peptides, MS identification in untreated control samples should be severely impaired by the presence of the nucleotide remnant crosslinked to the RBDpep.

Optionally, a third fraction can be collected after the first round of oligo(dT) capture and protease treatment. This sample is referred to as “input” and closely corresponds to the RNA interactome[10] with the addition of the LysC or ArgC proteolytic treatment.

Therefore, this control can serve a dual purpose: i) generate the repertoire of RBPs of the cell under study if not yet established, and ii) assess the influence of the protease of choice on the MS analysis. The input fraction is important in case of poor assignments of RNA-binding sites, since it can be used to test whether the first protease used is incompatible with the downstream tryptic digestion. Such an outcome will be reflected in the inefficient peptide identification in the input fraction because the peptides generated are too short.

We recommend controls with a non-irradiated sample (noCL) to define the background as in[10]. Peptide intensity ratios between UV-irradiated versus non-irradiated sample in three independent biological replicates are analyzed, applying statistics (e.g. moderated t-test) to remove false positives. We recommend using a 1% FDR to minimize the occurrence of false positives, and for consistency with other RNA interactome studies.

Number of cells. The modified protocol described here makes use of a substantially lower number of cells than in the original protocol[10]. Applied to HeLa cells, this workflow lead to the determination of an RNA interactome comparable in quantity and quality to the original study using 1/8 of starting material[9,29]. However, the number of cells required may vary when working with small cells or cells growing in suspension.

Optimization of the protocol. A simplified protocol (from Step 1 to 14) using small-scale settings can be used for optimization of the two key steps, UV crosslinking and oligo(dT) capture, to the cellular system or organism of interest. These set up experiments can be performed with only two 15 cm^2 dishes per condition, allowing the

analysis of multiple parallel conditions in one experiment. RBP and RNA isolation can be monitored using silver staining (Figure 2C), western blotting, RT-qPCR or bioanalyzer as in[10]. In small-scale experiments, RNA isolation is performed with 400 µl of bead slurry, using 1.8 ml of lysis and wash buffers and 200 µl of elution buffer per oligo(dT) capture round (replacing volumes from Step 1 to 14). Silver staining should yield a complex protein pattern applying cCL or PAR-CL, whereas non-irradiated samples are expected to be devoid of proteins. Known RBPs are expected to be enriched in irradiated samples and absent in non-irradiated controls[9,10]. Negative controls (e.g. ß-actin and histone 3 and 4) should be absent in all eluates irrespective of the UV irradiation. A successful experiment should lead to at least 10 fold enrichment of mRNA over rRNA by RT-qPCR measurements[9,10]. As a rule of thumb, the UV dosage that promotes the most efficient capture of RBPs without affecting RNA integrity is optimal.

The protocol described here is tailored to adherent cells. For suspension cells, we recommend using a modified protocol where first cells are collected by low speed centrifugation, followed by resuspension of the cell pellet in 1 ml of PBS. The cell suspension is then transferred to a 10 cm dish and homogeneously distributed across the dish surface. The dish is subsequently placed on ice without the lid and irradiated as in Step 3. Cells are then lysed by addition of 2x lysis buffer (40 mM Tris HCl pH 7.5, 1M LiCl, 1% LiDS (wt/v, stock 10%), 2 mM EDTA, 10 mM DTT). For fragile cells or cells with poor adherence, we recommend avoiding PBS washes prior to UV irradiation (Step 2). Instead, we obtained optimal RNA interactome and RBDmap results when we omitted phenol red from the standard DMEM medium (unpublished

data, AC). Under these conditions, cells can be directly irradiated with UV after removal of the media.

In the present protocol, we used both cCL and PAR-CL in parallel to maximize the identification of RNA-binding sites. However, users may favor one crosslinking approach over the other due to technical limitations of the model system (e.g. limited 4SU incorporation in complex organisms or non-dividing cells) or due to prior information regarding the protein of interest (e.g. the RBP crosslinks more efficiently either with cCL or PAR-CL[9]).

Choice of protease. Once the first part of the protocol is established in the model system of choice, the next step is to choose the protease to use in Step 16. In our initial study, we used ArgC and LysC because of their i) compatibility with the downstream trypsin treatment, ii) suitability for proper digestion of most RBPs (~17 aa peptide fragments on average) and iii) experimentally proven efficiency in MS studies. However, these proteases are only suitable to generate RBD profiles of a substantial subset of the RNA interactome[9,29]. When studying a particular RBP, we recommend analyzing its sequence to identify the most suitable protease for RBDmap. As a reference, the protease of choice should generate fragments of about 20 aa length on average and contain internal tryptic peptides. Generation of longer proteolytic fragments will affect the RBD resolution, while shorter fragments may negatively affect the MS identification, because the probability to encounter an optimal trypsin cleavage site within the fragment will be reduced. For example, we followed an *in silico* approach analyzing the protein sequences of the HeLa mRNA interactome[9] to generate a set of predicted in silico proteolytic fragments for each of the ten proteases commonly used in proteomics (Figure 2B) (Supplemental Methods, page 1). We then mapped the tryptic peptides experimentally identified in the HeLa RNA interactome to these

fragments as an estimate of the expected peptide coverage[9]. Each tryptic peptide was extended in silico to the closest protease cleavage sites to reconstitute the expected peptide fragments. We set an arbitrary RNA-binding site at the center of each protein and determined the number of proteins for which a given protease would cover the RBD with a peptide shorter than 20% of the actual protein length. This analysis revealed LysC and ArgC as the most suitable proteases for RBDmap experiments. Nevertheless, other proteases such as AspN, chymotrypsin and glutamyl endopeptidase also deserve consideration (Figure 2B). If interested in a particular protein or protein family, we recommend performing similar in silico or manual sequence analyses to determine the most suitable protease.

The next important step is to determine the optimal amount of protease required to minimize the incidence of miscleavages. Figure 2D shows a protease calibration experiment in which we divided the eluate of a RNA interactome capture from five 500 cm^2 dishes into five tubes and tested different LysC concentrations (from 0.5 to 3 µg) and incubation temperatures (37°C and 22°C). Incubation with 1 µg of LysC at 37°C is sufficient to fully disrupt the protein pattern observed in untreated cells, suggesting that proteins are efficiently proteolyzed. Indeed, no polypeptides (except the protease itself) are detected in a 15% acrylamide gel, indicating that the protease cleaves the RBPs into small peptides that run out the gel (Figure 2D). By contrast, treatment with 3 µg of LysC for 16h at 22°C does not promote complete cleavage of purified RBPs. MS analyses of RNA interactome capture eluates treated with the experimentally determined concentrations of LysC and ArgC revealed that both proteases generate peptides of about 17 amino acids on average and that the number of missed cleavages is negligible[29]. Importantly, RNA is massively degraded when the protease is not added to the digestion mix, suggesting the presence of contaminant RNases that

damage the sample in extended incubations at 37°C (Figure 2E). However, protease treatment prevents RNA degradation, probably by proteolytic cleavage of the RNases. We nevertheless add RNase inhibitors to prevent any potential RNA degradation that may still occur immediately after protease addition.

The two rounds of oligo(dT) capture. The first purification using oligo(dT) is performed with cell lysates and is designed to remove all non-covalent RNA interactors using high salt (500 mM LiCl) and ionic detergents (0.5% LiDS). In contrast, the second round of oligo(dT) capture is designed to separate the peptide pool that remains covalently bound to RNA after the first protease treatment from the pool of peptides that is released into the supernatant. Because the supernatant will be directly analyzed by MS to identify the released peptides, we recommend avoiding the use of MS-incompatible detergents such as LiDS or NP-40 (igepal) in the hybridization buffer. Subsequent washes containing high salt concentrations and the ionic detergent LiDS prevent co-purification of contaminant released peptides via non-covalent interactions with the RNA or with the peptides crosslinked to it. In addition, these washes increase the dilution factor, minimizing the contamination of the RNA-bound fraction with highly abundant released peptides. The second oligo(dT) capture step is performed in a significantly lower volume (2 ml vs 25 ml) to facilitate the downstream sample preparation for MS and to maximize the capture of the RNA. We thus adjusted the amount of beads in the second oligo d(T) selection step to maintain the volume to beads ratio.

Protease digestion. To optimize the protease treatment, we recommend incubating the eluates from Step 14 with increasing concentrations of protease and selecting the minimal concentration to fully disrupt the protein pattern observed by silver staining (Figure 2D). As a necessary control for RBDmap assays, we recommend to collect aliquots prior to and after protease treatment to monitor protein and RNA integrity by silver staining and bioanalyzer (Figure 2D-G). The sample should be used for the second oligo(dT) capture only if the protease digestion appears complete. We noticed extensive RNA degradation when incubating the samples for 8 hours at 37°C in absence of protease, suggesting the presence of contaminating RNases (Figure 2E). We thus recommend supplementing the extracts with RNase inhibitors (e.g. RNase-In) to protect the RNA from degradation during the early steps of the protease treatment, when RNases are still intact. After longer incubations both the RNases and the RNase inhibitor will be hydrolyzed by the protease.

Mass spectrometry. We used dimethyl labeling[40] for quantitative proteomic analysis, employing a platform coupling nano-flow UPLC to a mass spectrometer with MS/MS capability, such as the quadrupole-time-of-flight or Orbitrap instrument (e.g., similar to[41]). To maximize the depth of peptide identification and reduce sample complexity, we used isoelectric focusing pre-analytically to generate 12 peptide fractions. Note that the number of fractions can be adjusted according to sample complexity and the total amount of peptides available. Each fraction is then analyzed by high-resolution nano-LC-MS/MS in data-dependent acquisition mode. A critical point in RBDmap is its quantitative power, and we successfully applied stable-isotope dimethyl labeling to quantify the peptides in input samples as well as in the RNA-bound and released fractions[29,30]. Note that this protocol is also compatible with other chemical labeling strategies, such as iTRAQ[42] or TMT[43]. In principle, label-free quantification is also

capable to provide quantitative information on peptide abundance, however, we recommend to apply a chemical labeling strategy to avoid problems with intensity normalization between RNA-bound and released samples.

HPLC and MS parameters. Typically, 90-120 min linear gradients from 5-35% solvent B are sufficient for optimal peptide separation. A longer 4h gradient would be advisable in cases where no peptide fractionation is performed. Other MS/MS-parameters are instrument- and vendor-specific and should be optimized for the maximum number of peptide identifications. We found that the standard settings commonly used in proteomic applications provide a good starting point, and usually allow satisfactory analyses. Mass spectrometric raw data can be analyzed with any standard free (e.g. MaxQuant[44]) or commercial software tools that allow peptide identification and provide peptide-level quantitative data.

MS data analysis. We processed the raw data using MaxQuant (version 1.3.0.5)[44]. MS/MS spectra should be searched against the appropriate UniProt database linked to a database containing common contaminants. For identification, we recommend to set the enzyme specificity to trypsin/P, allowing a maximum of two missed cleavages. We allowed cysteine carbamidomethylation as fixed modification and methionine oxidation and protein N-terminal acetylation as variable modifications, while the minimal peptide length was set to six amino acids. The mass tolerances were set to 20 ppm for the first search, 6 ppm for the main search, and 0.5 Da for product ion masses. We applied a 1% FDR for both peptide and protein identification.

Computational assignment of RBDs. Peptides identified by MS can be analysed using the statistical programming language R and the packages RBDmap, RBDmapHeLa, and mRNAinteractomeHeLa described in[9,29]. A detailed analytical

pipeline is provided in Supplementary Methods. Briefly, a simulation of tryptic peptides can optionally be performed to assess suitability of different proteases for the protein/s of interest (Supplementary Methods, section 2, pages 1-5). Peptide identifications and quantifications by MaxQuant are first mapped to a reference human proteome annotation that includes information on protein and gene identifiers, protein domains, and other regions (Supplementary Methods, sections 3-4, pages 6-10). Proteins enriched in the UV-irradiated over the non-irradiated input sample with a 1% FDR are classified as RBPs. To identify the RNA binding regions, the ratio between the intensity of each identified peptide in the RNA-bound and released fraction is computed (Supplementary Methods, section 5, pages 10-11). Statistical significance is determined by a moderated t-test corrected for multiple testing with Benjamini-Hochberg (Supplementary Methods, section 6, pages 11-16). Peptides enriched in the RNA-bound over the released fractions with 1% FDR are classified as RBDpeps and those enriched with a 10% FDR are catalogued as “CandidateRBDpep” (i.e. potential RNA-binding region). Peptides are then extended to their adjacent LysC and ArgC cleavage sites to recall the original proteolytic fragments, only if the tryptic peptide is uniquely mapped to a single protein (Supplementary Methods, section 7, pages 16-17). Proteolytic fragments are displayed in individual protein plots together with the domain information (Supplementary Methods, section 8, pages 17-19). Finally, an html report summarizing the analysis is generated (Supplementary Methods, section 9, pages 19-20).

Reagents

- 4-Thiouridine (4SU) (Sigma-Aldrich, cat. no. T4509).

- Acetonitrile, HPLC grade (Merck Millipore, cat. no. 1000302500)
- Amicon Ultra® Centrifugal Filters (15 ml and 500 µl, 3KDa cut off, Millipore cat. no. UFC900324 and UFC500396). CAUTION! Equilibrate the filter unit following the manufacturer's instructions before use.
- Calcium chloride (CaCl2) (Sigma-Aldrich, C5670).
- Complete EDTA-free proteinase inhibitor cocktail (Roche, cat. no. 11873580001).
- DMEM with 4.5 mg l^{-1} D-glucose (Sigma-Aldrich, cat. no. D0422).
- Endoproteinase ArgC Sequencing Grade (Promega, cat. no. V1881)
- Endoproteinase Lys-C (Wako, cat. no. 129-02541).
- Ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich, cat. no. E6758).
- Dimethylsulfoxide (DMSO) (Merck Millipore, 102952).
- Di-sodium hydrogen phosphate (Na2HPO4) (Merck, cat. no. 1.06580).
- Dithiothreitol (DTT) (Biorad, 161-0611).
- Formaldehyde (CH2O) (37% (vol/vol), (Sigma-Aldrich, cat. no. 252549).
  ! CAUTION. Formaldehyde solutions and formaldehyde vapors are toxic, prepare solutions in a fume hood.
- Formaldehyde (CD2O) (20%, 98% D, Isotec, cat. no. 492620).
  ! CAUTION. Formaldehyde solutions and formaldehyde vapors are toxic, prepare solutions in a fume hood.
- Formaldehyde (13CD2O) (20%, 99% 13C, 98% D, Isotec, cat. no. 596388).
  ! CAUTION. Formaldehyde solutions and formaldehyde vapors are toxic, prepare solutions in a fume hood.
- Formic acid, UHPLC grade (Sigma-Aldrich, cat. no. 14265).
- Glutamine (Gibco, cat. no. G7513).

- HeLa (American type culture collection (ATCC), cat. no. CCL-2). Maintain the cells at 37°C in a humidified atmosphere containing 5% CO2.

  CAUTION: The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.
- Heat inactivated fetal bovine serum (Gibco, cat. no. 10500064).
- Huh-7 cells (ATCC cat. no. HB-8064). Maintain the cells at 37°C in a humidified atmosphere containing 5% CO2.

  CAUTION: The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.
- Iodoacetamide (Biorad, cat. no. 163-2109).
- Lithium chloride (LiCl) (Sigma, cat. no. 62476).
- Lithium dodecyl sulfate (LiDS) (Sigma-Aldrich, cat. no. L9781).
- Oligo (dT_{25}) magnetic beads (New England Biolabs, cat. no. S1419S).
- Penicillin-streptomycin (Sigma-Aldrich, cat. no. P0781).
- Phosphate-buffered saline (PBS) tablet (Sigma-Aldrich, cat. no. P4417).
- Ribonuclease A from bovine pancreas (RNase A) (Sigma-Aldrich, cat. no. R4642).
- Ribonuclease T1 from *Aspergillus oryzae* (RNase T1) (Sigma-Aldrich, cat. no. R1003).
- RNase-In (Promega, cat. no. N2511).
- SilverQuest™, kit for silver staining (Invitrogen, cat. no. LC6070).
- Sodium chloride (NaCL) (Sigma-Aldrich, cat. no. S7653).
- Sodium cyanoborohydride (NaBH3CN) (Fluka, cat. no. 71435).
- Sodium cyanoborodeuteride (NaBD3CN) (96% D, Isotec, cat. no. 190020).
- Sodium dihydrogen phosphate (NaH2PO4) (Merck, cat. no. 1.06346).
- StageTips, C18 material, 200µL tip (Thermo Fisher Scientific, cat. No. SP301).

- Triethylammoniumbicarbonate (TEAB) (Sigma-Aldrich, cat. no. T7408).
- Trypsin Gold, Mass Spectrometry Grade (Promega, cat. no. V5280).
- Urea (Biorad, cat. no. 161-0745-MSDS).

Equipment

- 500 mm^2 square dishes (Nunclon, cat. no. 166508).
- Class II biosafety cabinet.
- Crosslinking devices: cCL = 254 nm bulbs; PAR-CL= 365 nm bulbs. Spectrolinker UV Crosslinkers (Spectroline cat. no. XL-1500 and XL-1500A, respectively).
- GP Millipore express plus 500 ml funnel (Millipore, cat. no. SCGPT05RE)
- Humidified 37°C, 5% CO_2 incubator.
- Magnetic Separation Rack (50 ml NEB, cat. no. S1507S; 15 ml, Permagen; 2 ml, NEB cat. no. S1509S or DynaMag-2, cat. no. 123.21D, Invitrogen).
- Nano-UPLC system (e.g. EASY-LC 1000 or similar; Waters nanoAcquity or similar), capable of operating at a flow rate of 100-300 nl/min at a pressure of up to 1000 bar.
- Needle (27G, 3/4-inch; no. 20, 0.4 mm × 19 mm; BD Microlance; cat. no. 302200).
- Refrigerated Bench-top centrifuge.
- RNA 6000 Pico Bioanalyzer Chip and reagents (Agilent technologies, cat. no. 5067-1513 and 5067-151).
- Software for protein identification by database searching, e.g. Mascot (MatrixScience), Sequest (Thermo) or Andromeda[45] (via MaxQuant[44]).
- Syringes (5 ml; luer lock; Medicine, ref: IVL05).

- Tandem Mass Spectrometer (e.g. Thermo Scientific Orbitrap Fusion, Q-Exactive or similar; Bruker Impact II or similar).
- MaxQuant software[44] (or similar) and R/bioconductor package limma[46]. A detailed analysis pipeline can be found in the Supplementary methods.
- RBDmap, RBDmapHeLa, and mRNAinteractomeHeLa R packeges[9,29]

Reagent setup

- (1M) DTT. Divide the stock of (1M) DTT into 5 ml aliquots and store at -20°C for 1 year.

  CRITICAL: DTT is used to minimize the contamination of indirect RNA interactors forming disulfide bridges with the actual RBP.
- 4-Thiouridine. Prepare 1 ml of a 100 mM stock solution in water and divide it into 50 µl aliquots. Protect the solution from light by covering the tube with aluminum foil and storing the aliquots into an opaque box at -20°C for 1 year.
- Oligo(dT) magnetic beads equilibration and recycling. To equilibrate the beads, wash them three times with 3 volumes of lysis buffer and incubate for 5 min with gentle rotation between washes. To recycle the beads between capture rounds, add 1ml of 0.1 NaOH to the pellet. Resuspend the beads vigorously and incubate for 3 min at 55°C with agitation in a thermomixer. Collect the bead pellet with a magnet and wash 3 times with 1.5 ml of PBS supplemented with 0.025% NP-40 prior to adding them again to the lysate. Beads can be recycled up to three times.
- Lysis buffer. 20 mM Tris HCl pH 7.5, 500 mM LiCl, 0.5% LiDS (wt/vol, stock 10%), 1 mM EDTA, 5 mM DTT. Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. After cooling, complete

the buffer by addition of filtered LiDS. DTT should be added immediately before usage. Can be stored for up to 3 months at 4°C.

CRITICAL: We typically generate 1l and then aliquot into 200 ml bottles (equivalent to the volume required for one large-scale experiment). By doing so, we noticed a reduction in contaminations due to handling (e.g. RNases, proteases, keratins, etc). CRITICAL: We do not observe traces of RNA degradation when combining EDTA and the denaturing agent LiDS in the lysis buffer.

- Buffer 1. 20 mM pH 7.5 Tris HCl, 500 mM LiCl, 0.1% LiDS (wt/vol), 1 mM EDTA, 5 mM DTT. Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. After cooling, complete the buffer by addition of filtered LiDS. DTT should be added immediately before usage. Can be stored for up to 3 months at 4°C. CRITICAL: We typically generate 1l and then aliquot into 200 ml bottles (equivalent to the volume required for one large-scale experiment). By doing so, we noticed a reduction in contaminations due to handling (e.g. RNases, proteases, keratins, etc).
- Buffer 2. 20 mM pH 7.5 Tris HCl, 500 mM LiCl, 1 mM EDTA, 5 mM DTT, 0.01% NP40/Igepal (vol/vol). Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. DTT should be added immediately before usage. Can be stored for up to 3 months at 4°C. CRITICAL: We typically generate 1l and then aliquot into 200 ml bottles (equivalent to the volume required for one large-scale experiment). By doing so, we noticed a reduction in contaminations due to handling (e.g. RNases, proteases, keratins, etc).
- Buffer 3. 20 mM pH 7.5 Tris HCl, 200 mM LiCl, 1 mM EDTA, 5 mM DTT. Autoclave to inactivate contaminants such as proteases that may potentially

interfere with RBDmap. DTT should be added immediately before usage. Can be stored for up to 3 months at 4°C. CRITICAL: We typically generate 1l and then aliquot into 200 ml bottles (equivalent to the volume required for one large-scale experiment). By doing so, we noticed a reduction in contaminations due to handling (e.g. RNases, proteases, keratins, etc).

- Elution buffer. 20 mM pH and 7.5 Tris HCl. Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. Can be stored for up to 3 months at 4°C.
- 5x Hybridization buffer. 100 mM Tris HCl pH 7.5, 2.5 M LiCl, 1 mM EDTA, 25 mM DTT. Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. DTT should be added immediately before usage. Can be stored for up to 3 months at 4°C. CRITICAL: We typically generate 1l and then aliquot into 200 ml bottles (equivalent to the volume required for one large-scale experiment). By doing so, we noticed a reduction in contaminations due to handling (e.g. RNases, proteases, keratins, etc).
- Buffer 4. 10 mM pH 7.5 Tris HCl, 50 mM NaCl. Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. Can be stored for up to one year at room temperature (22 to 25°C).
- 5X proteinase K buffer. 50 mM pH 7.5 Tris HCl, 750 mM NaCl, 1% SDS (wt/vol), 50 mM EDTA, 2.5 mM DTT, 25 mM CaCl2. Autoclave to inactivate contaminants such as proteases that may potentially interfere with RBDmap. DTT should be added immediately before usage. Can be stored for up to one year at room temperature (22 to 25°C).
- 10X ArgC buffer. 8M urea, 250 mM 7.5 Tris HCl, 25 mM $CaCl_2$, 25 mM DTT. Autoclave to inactivate contaminants such as proteases that may potentially

interfere with RBDmap. DTT should be added immediately before usage. Use freshly made.

CRITICAL: Always prepare urea-based buffers (ArgC buffer) freshly on the day of use. Aqueous urea can degrade to isocyanic acid which can react with proteins, hampering data analysis.

Mass spectrometry buffers

- 50 mM TEAB. 500 µL of 1M Triethylammoniumbicarbonate in 10 mL water. Store at 4°C for 1 year.
- 1M DTT. 154 mg Dithiothreitol in 1 ml water.

CRITICAL: Can be aliquoted and stored at -20°C for 1 year.

- 1M IAA. 185 mg Iodoacetamide in 1 ml water.

CRITICAL: prepare freshly and use within a day. Protect from light.

- 0.5 M NaCl. 29.2 mg Sodium chloride in 1 mL water. Store at room temperature (RT) indefinitely.
- 50 mM NaH_2PO_4: 6.9 g in 1 l water. Store at 4°C for 1 year.
- 50 mM Na_2HPO_4: 8.9 g in 1 l water. Store at 4°C for 1 year.
- 50 mM sodium phosphate buffer pH 7.5 (for 9 ml mix 2 mlof 50 mM NaH2PO4 with 7 ml of 50 mM Na2HPO4). Store at 4°C for 1 year.
- *light* labeling reagent: for 1 ml mix 900 µl 50 mM sodium phosphate buffer with 50 µl 4% (vol/vol) formaldehyde in water and 50 µl 0.6 M cyanoborohydride (NaBH3CN). Prepare freshly on the day of use and keep at 4°C until use, do not store for more than 24h.

- *medium* labeling reagent: for 1 mL mix 900 µl 50 mM sodium phosphate buffer with 50 µl 4% (vol/vol) D_2-formaldehyde (CD_2O) in water and 50 µl 0.6 M cyanoborohydride (NaBH3CN). Prepare freshly on the day of use and keep at 4°C until use, do not store for more than 24h.
- *heavy* labeling reagent: for 1 ml mix 900 µl 50 mM sodium phosphate buffer with 50 µl 4% (vol/vol) $^{13}C,D_2$-formaldehyde ($^{13}CD_2O$) in water with 50 µl 0.6 M cyanoborodeuteride ($NaBD_3CN$). Prepare freshly on the day of use and keep at 4°C until use, do not store for more than 24h.
- Buffer A: 1% (vol/vol) formic acid in water. Store at room temperature (RT) indefinitely.
- Buffer B: 1% (vol/vol) formic acid in 80% acetonitrile. Store at room temperature (RT) indefinitely.
- MS-Buffer: 5% (vol/vol) DMSO, 1% (vol/vol) formic acid in water. Store at room temperature (RT) indefinitely.

PROCEDURE

Cell culture TIMING overnight

1. Seed HeLa cells on 8 x 500 cm^2 (4000 cm^2 of total growth area) dishes at a confluence of 40%. Incubate the cells overnight in normal medium (DMEM and 5% fetal calf serum, FCS) to reach 80% confluence after 16 h (~$2x10^7$ cells per dish). Label half of the dishes with 4SU (i.e. four) for PAR-CL and keep the other half unlabeled (i.e. four) (Figure 3). To label the RNAs with the photoactivatable nucleotide analogue for PAR-CL, supplement the medium with 100 µM 4SU and incubate overnight. Separate one 4SU-labeled and one unlabeled dish to be used as non-irradiated negative control. Use three 4SU-

labeled and three unlabeled dishes for PAR-CL and cCL, respectively (Figure 3).

? TROUBLESHOOTING

CAUTION The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

CRITICAL STEP: (Optional) To test 4SU incorporation, a small scale (see adaptation) experiment using only PAR-CL is performed. 4SU-labeled poly(A) RNA present in oligo(dT) eluates is purified via 4SU-dependent biotinylation followed by streptavidin purification as described elsewhere[47,48]. Proportion of labeled RNA will be defined by comparing the amount of RNA in the eluates of the streptavidin purification with that remaining in the supernatant.

Preparation for UV-crosslinking TIMING ~30 min

2. Wash cells twice with 15 ml of PBS (room temperature, RT).

CRITICAL STEP: we noticed that residual DMEM can induce variability in crosslinking efficiency across experiments. We recommend removing carefully the excess of DMEM by tilting the dish and handling the 4SU-labeled non-irradiated control in a darkened room to avoid potential crosslinking mediated by natural light.

UV-crosslinking TIMING ~30 min (whole sample set)

3. Remove the PBS and place the culture dishes after removing their lids on ice immediately before crosslinking at a distance of ~15-20 cm from the UV source. Irradiate with 0.15 J/cm^2 (~25-50 sec) with UV light at 254 nm for cCL or at 365 nm for 4SU-labeled cells (PAR-CL).

? TROUBLESHOOTING

CRITICAL STEP: Pre-heat the UV bulbs (both 365 nm and 254 nm) by running a 0.5 J/cm^2 program immediately before starting the irradiation of the dishes. This step will make the intensity of the UV bulbs comparable across irradiation steps (i.e. similar J/cm^2 per second).

CRITICAL STEP: It is important to process the cells quickly from Step 3 to Step 5; therefore, if dealing with a large number of dishes, proceed with small sub-sets (about 1-3 dishes) at a time, while leaving the remaining dishes in culture. Non-irradiated controls should be processed last and kept in the incubator while working with the dishes treated with UV light.

Lysis TIMING ~15 min

4. For RNA bound/released fractions (two dishes of cCL and two dishes of PAR-CL; Figure 3): Add 10 ml of lysis buffer to one plate irradiated with UV light at 254 nm and harvest the cells with a rubber policeman. Transfer the 10 ml of lysate to the other plate irradiated with UV at the same wavelength and harvest the cells. Transfer these 10 ml of lysate to a 50 ml Falcon tube. Add 10 ml of fresh lysis buffer to one dish irradiated with 365 nm UV light and harvest the cells. Transfer the 10 ml of lysate to the second dish irradiated at the same wavelength and harvest the cells. Add these 10 ml to the same 50ml Falcon tube and mix with a 25 ml pipet. Adjust the volume to 25 ml using fresh lysis buffer.

For the input fractions (one dish of noCL, one dish of 4SU-labeled noCL, one dish of cCL and one dish of PAR-CL; Figure 3). Add 10 ml of lysis buffer to a dish irradiated at 254 nm and harvest the cells. Transfer these 10 ml to the dish irradiated at 365nm and harvest the cells. Transfer the lysate to a 15 ml Falcon

tube and adjust the volume to 12 ml. Repeat the same process with the dishes that were not UV-exposed, transferring the lysate to a new tube.

At this stage we have 1x50 ml tube with 25 ml of lysate for the RNA-bound and released fractions, and 2x15 ml tubes with 12 ml of lysate from irradiated and non-irradiated cells for the input fractions.

Homogenization TIMING ~30 min

5. Pass each sample three times through a syringe with a narrow needle (gauge 0.4mm diameter) to homogenize. Incubate the lysates for 10 min at 4°C.

CRITICAL STEP The bigger the syringe, the more difficult it is to pass the lysate through the narrow needle. We recommend using 5 ml syringes and transferring the lysate from one tube to a new one stepwise until the homogenization has been completed. The lysate should be passed through the needle quickly to be appropriately homogenized. If the sample is not completely homogeneous after 3 cycles, repeat as many times as needed, and do not proceed to the next step until the homogenization is satisfactory (i.e. viscosity is significantly reduced when inverting the tube and the sample looks homogeneous).

? TROUBLESHOOTING

PAUSE POINT Samples can be frozen at -80°C for up to one month. Nevertheless, we recommend avoiding repeated freeze-thawing cycles. If possible, proceed to Step 6 without freezing.

First oligo(dT) capture TIMING ~6h

6. Add 3 ml and 1.5 ml of previously equilibrated oligo(dT)$_{25}$ magnetic beads (see Reagent Setup) to the 50 ml and 15 ml falcon tubes, respectively. Incubate for 1 h at 4°C with gentle rotation. Place the tubes onto a magnet at 4°C and wait until the beads are completely captured. This process should take 2-10 min. Incomplete homogenization may explain longer times. Recover the supernatants and store them in new tubes at 4°C for the other two cycles of oligo(dT) capture (see Step 13).

7. Resuspend the beads in 25 ml and 12 ml of ice-cold lysis buffer for the 50 ml and the 15 ml tubes, respectively. Incubate for 5 min at 4°C with gentle rotation, pellet the beads with the help of a magnet. Remove and discard the supernatant.

CRITICAL STEP: We recommend resuspending the beads first in a small volume (4 and 2 ml of lysis buffer for the 50 ml and the 15 ml tubes, respectively) by vigorously pipetting up and down using a p1000 pipet. When satisfied with the resuspension, add lysis buffer to reach 25 ml and 12 ml, respectively. Repeat this process in Steps 8-10.

8. Add 25 ml and 12 ml of ice-cold buffer 1 for the 50 ml and 15 ml tubes, respectively. Resuspend the beads and incubate for 5 min at 4°C with gentle rotation. Pellet the beads using the magnet and discard the supernatant. Repeat this step once (i.e. two washes with buffer 1).

CRITICAL STEP: Note that when the mRNA-protein complexes are efficiently isolated, a halo surrounding the bead pellet will be noticeable during the first or second wash with this buffer only in the UV crosslinked samples[10].

9. Add 25 ml and 12 ml of ice-cold buffer 2 to the 50 ml and 15 ml tubes, respectively. Resuspend the beads and incubate for 5 min at 4°C with gentle

rotation. Pellet the beads using the magnet and discard the supernatant. Repeat this step once (i.e. two washes with buffer 2).

10. Add 25 ml and 12 ml of buffer 3 at room temperature to the 50 ml and 15 ml tubes, respectively. Resuspend the beads vigorously. Pellet the beads using the magnet and discard the supernatant. Repeat this step once (i.e. two consecutive washes with buffer 3).

CRITICAL STEP: We noticed bead loss occurring at this step following incubation with buffer 3 for too long. We thus recommend avoiding the 5 min incubation performed in the previous steps and placing the tube in the magnet immediately after resuspending the beads.

11. For elution, resuspend the beads in 500 µl or 250 µl of elution buffer for the 50 ml and 15 ml tubes, respectively, and transfer the bead suspensions to a low-binding, 2 ml tube. Incubate for 3 min at 55°C with agitation (~500 rpm in a thermomixer). Pellet the beads using the magnet and transfer the supernatant to a new tube. Beads can be recycled at this point as indicated in Reagent Setup. Determine the RNA concentration in eluates using a Nanodrop device.

CRITICAL STEP: (Optional) To ensure that the eluate is fully depleted of magnetic beads, we recommend placing the eluates on the magnet for 3 min at 4°C. Collect the supernatant again and transfer it to a new tube prior to measuring the RNA concentration in a Nanodrop device.

12. Repeat Steps 7 to 11 twice more to deplete samples of poly(A) RNAs. We recommend reusing the oligo(dT) beads up to three times – to do so, treat the beads as described in Reagent Setup. Then add them back to the lysate separated in Step 7.

CRITICAL STEP. Do not re-use the beads from one “condition” for other conditions (e.g. beads used for crosslinked samples should not be used for non-irradiated controls) to avoid cross-contamination.

13. Pool the eluates from the three successive cycles of poly(A)$^+$ RNA capture (final volume 1.5 ml for RNA bound/release fractions and 0.75 ml for the input fraction).

PAUSE POINT, samples can be stored at -80°C for up to one week.

Protease digestion TIMING ~9 h

14. Thaw the eluates from Step 13 at room temperature and collect before the proteolytic treatment 50 µl from each tube for protein analysis and 10 µl for RNA analysis.

15. Treat the eluates with either LysC (option A) or ArgC (option B) proteases.

A) LysC treatment

i) Add 1 µl of RNase-in and 3 µg of LysC to the RNA-bound/released fraction (~300 µg of eluted RNA in ~1.4 ml) and 0.5 µl of RNase-in and 1.5 µg of LysC to input samples (~150 µg of eluted RNA in ~690 µl) remaining after Step 14.

Ii) Incubate for 8h at 37°C with gentle agitation (~200 rpm).

B) ArgC treatment

i) Add 150 µl 10x ArgC buffer to the RNA-bound/released fraction (~300 µg of eluted RNA in ~1.4 ml) and 75 µl to the input sample (~150 µg of eluted RNA in ~690 µl) remaining from Step 14 to obtain a ~1x solution. Supplement the RNA-bound/released fraction with 1 µl of RNase-in and 1 µg of ArgC and the input fraction with 0.5 µl RNase-in and 0.5 µg of ArgC.

iI) Incubate for 8h at 37°C with gentle agitation (~200 rpm).

CRITICAL STEP: (Optional) The protocol is described for the use of either LysC or ArgC, which are optimal for a large proportion of human RBPs. However, the use of other proteases may be required (see Experimental Design). In this case, adapt the digestion buffer accordingly to the manufacturer's recommendations, avoiding the use of buffers with basic pH (higher than 8) to prevent alkaline hydrolysis of the RNA and other regents that may harm the RNA or the protein.

? TROUBLESHOOTING

16. After the proteolytic treatment, collect 50 µl and 10 µl for protein and RNA analyses, respectively.

PAUSE POINT, samples can be stored at -80°C up to one week.

17. Supplement the input samples (~700 µl) derived from the oligo(dT) capture in 15 ml tubes with 50-10U of RNase T1 and RNase A and incubate for 1 h at 37°C. After RNase treatment input samples can be processed for MS.

CRITICAL STEP: We noticed that this RNase treatment can efficiently degrade the RNA even without inactivation of the protease. This suggest that most of the protease is lost after the 8h of incubation. Thus, no inactivation step is required.

? TROUBLESHOOTING

Protein and RNA analyses – TIMING ~1day

18. Analyze the proteolytic treatment and RNA integrity by silver staining (following the manufacturers' recommendations) and bioanalyzer[9]**,** respectively, using the 50 µl and 10 µl fractions collected before and after the proteolytic digestion (in Steps 14 and 16, respectively).

CRITICAL STEP: Prior to the protease digestion, silver staining should reveal a complex protein pattern distributed all across the molecular sizes (Figure 2F and G). Since the protease cleaves every 17 amino acids in average, proteolytic treatment should lead to a gel lane devoid of proteins, with the exception of the protease. In contrast, RNA is expected to show the same molecular size distribution in the bioanalyzer profile in both non-treated and treated samples (Figure 2E). Only when both are satisfactory, proceed to the next step.

Second oligo(dT) capture TIMING ~5h

19. Use the remaining sample of RNA-bound/released fractions from Step 16 (~1.4 ml; i.e. 500 µl from each of the three rounds of oligo(dT) capture minus the samples taken for quality controls in Steps 14-18) and add 10x hybridization buffer (~150 µl) to reach ~1x hybridization buffer.

? TROUBLESHOOTING

20. Add 500 µl of pre-equilibrated oligo(dT) magnetic beads (see Reagent Setup) to the sample and incubate for 1 h at 4°C with gentle rotation.

21. Pellet the beads by placing the tube on the magnet. Wait until the beads are fully collected on the magnet. Next collect the supernatant, transfer it into a new 2 ml tube and keep on ice until Step 28.

22. Resuspend the bead pellet vigorously in 1 ml of ice-cold lysis buffer by pipetting up and down. Transfer the bead suspension into a 15 ml falcon tube. Fill with 4 ml of lysis buffer. Incubate for 5 min at 4°C with gentle rotation. Next, collect the beads using the appropriate magnet rack and discard the supernatant.

23. Resuspend the pellet with 1 ml of ice-cold buffer 1 by vigorously pipetting up and down. Add 4 ml of buffer 1 and mix with the pipet. Collect the beads with the magnet rack and discard the supernatant.

24. Resuspend the pellet with 1 ml of ice-cold buffer 2 by pipetting vigorously up and down. Add 4 ml of buffer 2 and mix with the pipet. Collect the beads with the magnet rack and discard the supernatant.

25. Resuspend the pellet in 1 ml of buffer 3 (room temperature) by vigorously pipetting up and down. Add 4 ml of buffer 1 and mix with the pipet. Collect the beads with the magnet rack and discard the supernatant.

26. Resuspend the bead pellet in 250 µl of elution buffer (room temperature) and transfer the bead suspension into a new low-binding, 1.5 ml tube. Incubate for 5 min at 55°C with agitation in a thermomixer (~500 rpm). Move the tube to the magnet to pellet the beads and transfer the supernatant into a new low-binding, 1.5 ml tube. To ensure that no residual beads remain in the eluate, put the tube in the magnet again and transfer the supernatant into a new low-binding, 1.5 ml tube. Measure the RNA content (nanodrop) and keep the tube on ice.

27. Repeat Steps 21 to 26 twice using recycled beads (see Reagent Setup). Combine the eluates from the three 'second oligo(dT) capture' cycles and take 10 µl for RNA analysis (bioanalyzer).

28. Collect the supernatant (released fraction) from Step 21 and the pooled eluate (RNA-bound fraction) from Step 27 and treat them with 10U of RNase A and T1 for 30 min at 37°C.

? TROUBLESHOOTING

Sample preparation for MS - TIMING ~2 days

29. Adjust both the RNA-bound and released peptide solutions from Step 28 to 5 mM DTT by adding the appropriate amount of 1M DTT. Incubate the samples for 30 min at 56°C.

30. Add 1 M Iodoacetamide to a final concentration of 10 mM, and incubate the samples for 30 min at 20°C in the dark.

31. Transfer the samples into 3 kDa centrifugal filters. Adjust the volume to 400 µl with 50 mM TEAB and concentrate the sample by centrifugation at 14,000 x g for 30 min at 4°C.

CRITICAL STEP Applicable to steps 31 to 37. The samples are buffer-exchanged, concentrated, and digested using Amicon Ultra centrifugal filters (0.5 ml, 3-kDa cutoff) according to the FASP protocol[49] with minor modifications as indicated. Avoid primary amines in all buffers as they interfere with stable-isotope dimethyl labeling.

? TROUBLESHOOTING

32. For buffer exchange, add 300 µl of 50 mM TEAB and centrifuge at 14,000 x g for 30 min at 4°C. Discard the flowthrough.

33. Add sequencing grade trypsin at an enzyme-protein ratio of 1:50 to the samples. Since the absolute amount of protein in the samples is unknown at this point, add 0.5 µg trypsin in 50 µl 50 mM TEAB to the filter unit. Make sure the lid of the filter unit is well closed and then mix by vortexing. Insert the tube in a floating rack and incubate the filter in a wet chamber (e.g. a pipette tip box containing few ml of water; it is not necessary to seal the box) at 37°C for 14 hours.

? TROUBLESHOOTING

CRITICAL STEP: (optional) To ensure that the X-link peptides actually form part of the in silico generated RBDpep (LysC/ArgC fragment), an optional control omitting the treatment with trypsin can be performed (see Experimental Design).

34. Place the filter on a new 2 ml tube. Collect the peptides by centrifugation at 14,000 x g for 30 min at room temperature. Tryptic peptides will be present in the flowthrough (i.e. take the flowthrough).

35. Wash the filter membrane by adding 50 µl of 0.5 M NaCl. Centrifuge at 14,000 x g for 20 min at RT. Mix this flowthrough with that from Step 34.

36. Invert the filter unit and recover the remaining sample by centrifugation at 1,000 x g for 1 min at RT. Combine this eluate with the flowthrough of Step 35.

37. Acidify the combined sample (eluate from Step 36) by adding formic acid to a final concentration of 1%.

PAUSE POINT Samples can be frozen indefinitely at -20°C or -80°C.

Stable-isotope peptide labeling and fractionation – TIMING 1 day

CRITICAL: Stable isotope labeling by reductive amination in Steps 38-46 is performed according to the protocol of Boersema *et al*[40]. All labeling steps are carried out at room temperature on StageTips[50]. For complete labeling of primary amines, labeling solutions need to be prepared freshly on the day of usage. Avoid using larger SPE cartridges such as Sep-Pak since the concentration of peptides in the samples is low, and sample losses would be significant.

38. Take two StageTips for the RNA-bound and released samples to separately label the peptides with *light* and *heavy* dimethyl labeling solutions, respectively.

If also analyzing the input samples, take two additional StageTips, one for the noCL control and one for the cCL sample. Input samples can be labeled with the same isotope pairs (light and heavy) since they will be analyzed as an independent experiment in a separate MS run.

CRITICAL STEP: (optional) the *medium* dimethyl labeling reagent is not used here and can be used to label peptides from a paralleled non-irradiated or trypsin omitted negative control; alternatively, it can be used for a RNA-bound sample generated under different physiological conditions (see Introduction).

39. To activate the StageTips load 20 µl of acetonitrile and centrifuge for 1 min at 500 x g at RT. Discard the flowthrough.

40. Wash the StageTips with 20 µL of buffer B, centrifuge for 1 min at 500 x g at RT and discard the flowthrough.

41. Equilibrate the StageTips with 20 µL of buffer A, centrifuge for 1 min at 1,000 x g at RT. Discard the flowthrough.

CRITICAL STEP Adjust the centrifugation times if the solutions have not fully passed through the StageTips. Avoid letting StageTips run completely dry during the equilibration, loading and washing steps.

42. Load the whole peptide sample from the RNA-bound and released samples from Step 37 on the two activated StageTips. In other words, one StageTip will be loaded with the peptides from the RNA-bound fraction and the other with the released peptides. Load in separated StageTips the noCL and cCL input samples. Centrifuge for 2 min at 1,000 x g at RT.

CRITICAL STEP: Load the sample in multiple steps if the volume exceeds the capacity of the StageTip. Increase the centrifugation time if necessary until the entire sample volume has passed through the StageTip.

43. Wash the StageTips with 20 µL of buffer A, centrifuge for 1 min at 1,000 x g at RT. Discard the flowthrough.

44. Add 50 µL of the respective light, medium or heavy labeling solution for one round of labeling on the StageTips. Centrifuge at 500 x g for 10 min at RT. Discard the flowthrough. Repeat this step twice.

? TROUBLESHOOTING

45. Elute the labelled peptides by adding 50 µL buffer B and centrifuge 5 min at 500 x g at RT. Combine the differentially labeled eluates into a new tube (i.e. combine RNA-bound and released peptides in one tube and the noCL and cCL input samples in other tube)**.**

CRITICAL STEP: (optional) Peptide fractionation by isoelectric focusing[51], ion-exchange[52] or basic pH reversed phase chromatography[53], can be performed at this point to reduce sample complexity. Fractionation increases the sample preparation time but, in general terms, will improve peptide identification in complex samples. All the fractionation methods indicated perform similarly, thus the selection of the approach should rely on the availability of equipment and expertise in the laboratory.

46. Dry the sample by vacuum centrifugation.

PAUSE POINT Samples can be frozen indefinitely at -20°C until mass spectrometric analysis.

Peptide identification and quantification by LC-MS/MS – TIMING 1-5 days

47. Reconstitute dried samples in 10µl of 5% DMSO and 1% formic acid.

48. Inject an appropriate amount of sample to optimally load the LC column. Apply a linear gradient of organic solvent for peptide separation using nano-flow

HPLC. Analyze the sample using data-dependent acquisition (see Experimental Design).

CRITICAL STEP: The optimal amount of peptides will depend of the efficiency of the isolation, sample preparation and MS equipment. We recommend a preliminary analysis with 10% of the sample and, based on the total peptide intensity, adjust the amount loaded to avoid overload or underload of the MS system.

49. Analyze the RAW data using a software that employs peptide identification by database searching[54]. Filter the peptide identifications to an FDR of 1%. Extract ion chromatograms for each peptide and labeling state and determine heavy-to-light ratios

Data analysis – TIMING ~1 Week

CRITICAL: A step by step documentation of the data analysis and the scripts used can be found in the Supplemental Methods.

50. Extract ion count measures for each peptide using e.g. MaxQuant software[44].

51. Analyze the peptide intensities with the software and proteome references implemented in R in packages RBDmap, RBDmapHeLa, and mRNAinteractomeHela[9,29] . These packages will be used from Step 51 to 55 as indicated in the Supplemental methods. Summarize the ion count ratios between RNA-bound and released peptides for the identification of RNA-binding sites, and between noCL and crosslinked input samples, using information from three biological replicates (see Supplemental Methods, sections 3-4). The RNA-bound/released peptide ratios will be used to determine the RNA-binding regions, while the cCL/noCL input peptide ratios will serve to

determine the RNA interactome and to analyze the impact of the protease of choice on the MS peptide identification (see Supplemental Methods, section 2).

52. Plot the density of ratios from RNA-bound over released peptides intensities. Identify the center of the low-intensity mode (representing RNA-released peptides) with a robust estimate. Normalize the value such that the center of the mode becomes zero (see Supplemental Methods, section 5).

CRITICAL STEP The distribution of ratios is expected to be bi-modal, with one mode representing the RNA bound peptides and the other mode the released ones.

53. Test the normalized RNA-bound/released ion count ratios against the hypothesis that ratios are zero using three or more replicates by a moderated t-test implemented in the R/bioconductor package limma[46] (see Supplemental Methods, section 6).

TROUBLESHOOTING

CRITICAL STEP: (optional) In our original RBDmap experiments[29,30] we did not notice a large proportion of “infinite” or “zero” ratios (i.e. situation occurring when a given peptide is not detected in one of the two fractions, resulting in a zero value in the numerator or denominator that impedes the generation of a ratio). However, if the incidence of these is high in a given experiment, a semiquantitative approach that considers the reproducibility of a “zero” or “infinite” peptide intensity ratio across biological replicates can be implemented as in[15,38].

54. Correct p-values by the method of Benjamini-Hochberg controlling for FDR. We apply a FDR of 1% for both identification of RBPs (cCL versus noCL input

samples) and determination of RNA-binding sites (RNA-bound versus released peptides, Figure 2A) (see Supplemental Methods, section 6).

55. Extend the tryptic peptide to the adjacent LysC or ArgC cleavage sites to recall the original LysC or ArgC fragments. Fragments enriched in the RNA-bound fraction with 1% FDR are classified as RBDpeps. FDR cutoff of 10% is further used to identify candidate RBD peptides (See Supplemental Methods, section 7).

TROUBLESHOOTING

Troubleshooting advice is summarized in Table 1.

TIMING

Note that the time required to carry out Steps 1-16 can be reduced by scaling down the experiment or by having more than one person involved in the experiment.

Step 1	**Cell culture**	overnight
Step 2	**Preparation for UV-crosslinking**	~30 min
Step 3	**UV-crosslinking**	~30 min
Step 4	**Lysis**	~15 min
Step 5	**Homogenization**	~30 min
Steps 6-13	**First oligo(dT) capture**	~ 6h
Steps 14-17	**Protease digestion**	~ 9h
Step 18	**Protein and RNA analyses**	~1 day

Steps 19-28	**Second oligo(dT) capture**	~5h
Steps 29-37	**Sample preparation for MS**	2 days
Steps 38-46	**Stable-isotope peptide labeling and fractionation**	1 day
Steps 47-49	**Peptide identification and quantification by LC-MS/MS**	1-5 days
Steps 50-55	**Data analyses**	~1 week

ANTICIPATED RESULTS

Analysis by mass spectrometry of the UV irradiated versus non-irradiated input samples (Figure 1) generated a protein dataset[29] that closely resembles the previously described human RNA interactome (82% overlap; 1% FDR)[8,9,13], implying that the LysC and ArgC treatments do not negatively affect protein identification by MS. Therefore, the experimental protocol described here also permits the determination of comprehensive RNA interactomes by analyzing the input fraction.

Analysis of the RNA-bound and released fractions (Figure 1) by MS is expected to reveal thousands of peptides. Peptide intensity ratios between the RNA-bound and released fractions should follow a bimodal distribution (Figure 2A), with one mode representing the released peptides (grey dots) and the other the RNA-bound ones (red dots). A problem separating the two fractions or a peptide loss due to inefficient sample preparation is likely to yield a unimodal distribution (see Troubleshooting). RNA-bound/released ratios should be consistent across the biological replicates (Figure 2A)[29]. Identified tryptic peptides are then extended to their adjacent LysC or ArgC cleavage sites to recapitulate the protein fragments generated by these proteases during the sample digestion. Proteolytic fragments enriched in the RNA-bound fraction

with 1% and 10% FDR are referred to as RBDpeps and candidateRBDpeps, respectively. For most of the analyses we focused on the 1% FDR data sets; however, a significant fraction of bona fide RNA-binding sites can be contained in the 10% FDR dataset, which are therefore also displayed in the individual protein profiles (Figure 4A and B).

Proteolytic fragments are plotted in the individual protein profiles taking their position within the protein into consideration, represented from N- to C-terminus on the x axes (Figure 4A and B). Domains are displayed as boxes below the x axis, facilitating the visual analysis of the RNA-binding sites against the domain distribution. The peptide 'fold change' between RNA-bound and released fractions is shown on the y axis. RBDpeps (peptides enriched in the RNA-bound fraction at 1% FDR) are represented in red, candidateRBDpep (peptides enriched in the RNA-bound fraction at 10% FDR) in salmon and released peptides (peptides enriched in the released fraction at 1% FDR) in blue. In HeLa cells, the position of RBDpeps strongly correlates with classical RBDs such as RRM type 1 ($p=1.7x10^{-79}$), RRM type 6 ($p=1.12x10^{-17}$), CSD ($p=9x10^{-5}$) and Zinc fingers CCHC ($1.4x10^{-3}$), showing that RBDmap efficiently identifies most of the well-established RNA-binding architectures (Figure 4A)[29]. Strikingly, RBDmap identified dozens of novel RBDs such as Ezrin/radixin/moesin (ERM, $p=2.45x10^{-4}$), high mobility group box domain (HMGB, $6.x10^{-4}$), heat shock protein 70 domain (HSP70, $p=8.9x10^{-3}$), PDZ domain (also known as Dlg homologous region, DHR, $p=9x10^{-3}$) and FK506-binding protein (FKBP) domain ($p=9x10^{-3}$), as well as more than six hundred RNA-binding sites mapping to intrinsically disordered regions[29]. These globular and intrinsically disordered RBDs are conserved across homologous and non-homologous proteins. RNA-bound fragments are more conserved across evolution than released regions, and are enriched for post-translational modifications,

suggesting biological significance[29]. For example, ERM proteins interact with the plasma membrane and cytoskeleton and were not previously associated with RNA biology. RBDmap identifies multiple LysC and ArgC RBDpeps overlapping at the C-terminus of the ERM domain of MSN and RDX (Figure 4B and C). Interestingly, recent RBDmap datasets from our lab generated from Huh-7 cells (unpublished, RH) and mouse HL-1 cardiomyocytes[30] show identical binding sites in these proteins, suggesting data consistency across biological systems and functional conservation from mouse to human (Figure 4B and C). The complete HeLa RBDmap dataset can be accessed at: http://www-huber.embl.de/users/befische/RBDmap/.

ACKNOWLEDGEMENTS

Our colleagues Bernd Fischer and Kathrin Eichelbaum unexpectedly passed away during the time in which this manuscript was revised. We dedicate this work to their memory. We thank Drs. Benedikt Beckmann and the Hentze group for helpful discussions. A.C. is funded by MRC Career Development Award #MR/L019434/1. M.W.H. acknowledges support by ERC Advanced Grant ERC-2011-ADG_20110310 and the Virtual Liver Network of the German Ministry for Science and Education. C.K.F. is supported by EMBO postdoctoral fellowship LTF1006-2013.

AUTHOR CONTRIBUTIONS

A.C., B.F., and M.W.H. conceived and designed the project. A.C., R.H., and A-M.A. carried out experimental work. C.K.F., S.F., and J.K. performed the proteomic analyses. B.F., A.I.J, T.C., A.C., C.K.F., J.K., and M.W.H. performed the data analyses. A.C. and M.W.H. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1: Schematic representation of RBDmap. RBP, RNA-binding protein. RBDpep refers to the RNA-bound peptide generated by LysC or ArgC proteolysis. RBDpeps are hydrolyzed by trypsin treatment into two type of peptides: the RNA-crosslinked peptide (X-link) and its neighboring tryptic peptides (N-link). N-link peptides exhibit normal mass and are used to identify the RBDpeps. Adapted with permission from[29].

Figure 2: RBDmap quality controls and data analysis. A) Scatter plot comparing the peptide intensity ratios between RNA-bound and released fractions from two independent experiments (i.e. replicates). 1%FDR RBDpeps and 10% FDR candidate RBDpeps are shown in red and salmon, respectively. (r: Pearson correlation coefficient). The scatter shows an excellent correlation between biological replicates (r= 0.85) and that the identified peptides follow a bimodal distribution. One mode represents the RNA-bound peptides (N-link) in red, and the other the released peptides (grey). B) In silico prediction of protease suitability for RBDmap (see section Anticipated results). The bar plot represents the number of proteins for which protease cleavage would lead to the identification of an RNA-binding site smaller than the 20% of the protein length (i.e. high resolution RBD identification). C) HeLa cells were incubated either in the presence or the absence of 4SU. Non 4SU-treated cells were UV_{254} irradiated (cCL) or mock-irradiated (noCL), while 4SU-labeled cells were UV_{365} irradiated (PAR-CL) or non-irradiated (4SU noCL). After oligo(dT) capture, proteins were analyzed by silver staining. A complex protein pattern (composed by RBPs) is observed only when cells are irradiated with UV light. In absence of irradiation, the

resulting lane is devoid of proteins. D) Optimization of LysC treatment. After UV treatment and oligo(dT) pulldown, eluates were treated with increasing amounts of LysC for 8h at 37°C or for 16h at 22°C. Protein integrity is analyzed by SDS-PAGE on a 15% acrylamide gel, followed by silver staining. The complex pattern of RBPs gradually disappears after treatment with increasing amounts of LysC. Complete protein cleavage (except for LysC itself**,** which is indicated with an asterisk (*)) is achieved with 1µg of LysC. E) RNA integrity from eluates in panel D was assessed using a bioanalyzer (RNA pico chip). After protease treatment, all samples show the same RNA pattern than the eluate from Step 13 (i.e. prior to the addition of the protease). However, incubation for 8 h at 37°C in absence of protease leads to a substantial increase in RNA products of low molecular size (lane 2), suggesting RNA degradation. F and G) Protein integrity quality controls of an RBDmap experiment in HeLa (F) and Huh-7 (G) cells. After UV crosslinking and oligo(dT) capture (cCL lanes), RBPs were eluted and treated with 3µg of LysC for 8h at 37°C. Samples taken before (0h LysC) and after (8h LysC) the protease treatment were analyzed by silver staining. Almost no protein is detected in the non-irradiated (noCL) control irrespective of the protease treatment. Conversely, a complex protein pattern is observed in the UV irradiated samples, which disappears after 8h of protease treatment at 37°C. The LysC and RNase T1 lanes show the mobility of these proteins, indicated by an asterisk, in a SDS-PAGE gel. Panels B, D and E are reprinted with permission from[29].

Figure 3. From cell culture to lysis. Workflow of the treatments performed from cell seeding until lysis, indicating the number of dishes used for each control and experimental sample. 4SU, 4-thiouridine; noCL, non-irradiated; cCL, conventional crosslinking (254 nm UV light); PAR-CL, photoactivatable ribonucleoside–enhanced crosslinking (i.e. 4SU mediated crosslinking at 365 nm UV light).

Figure 4. Analysis of individual RBDmap protein profiles. A) The RRM-containing proteins NCL and eIF3G are shown applying the single protein profile display. The X axis represents the protein length from N- to C-terminus, while the Y axis shows the fold change between RNA-bound and released fractions for each individual peptide. LysC and ArgC fragments are shown in red (1%FDR RBDpeps), salmon (10%FDR candidateRBDpeps) or blue (released fragments) lines. B) Individual map of the MSN protein from HeLa and Huh-7 RBDmap data. The protein profile is displayed as in (A). C) Homology model of the ERM domain of MSN using phyre2[55]. RBDpeps from HeLa (left) and Huh-7 (right) cells are displayed in red.

Supplementary Methods. RBDmap data analysis pipeline. A detailed example analysis workflow for RBDmap data based on R packages described in[9,29,30].

Table 1. Troubleshooting advice.

Step	Problem	Possible reason	Solution
49	Low peptide yield after MS	Low number of cells	Calibrate the number of cells by performing an experiment with a fixed amount of beads and lysis volume but increasing amount of cells. We noticed that more cells are required when their volume is small (e.g. mouse embryonic stem cells) or when they express pigments that interfere with UV irradiation (e.g. chlorophyll).
18	Inefficient crosslinking	Protocol not optimised for respective sample; e.g. UV-absorbing pigments	Some cell lines may require optimization (e.g. fragile and suspension cells or cells with small cytoplasms). We recommend testing different processing approaches and energies of UV irradiation if the crosslinking is inefficient compared to the same amount of total starting material in HeLa cells (protein in the whole cell lysate).

5	Viscous lysate	Chromatin precipitation	Initially, repeat the homogenization cycles until the viscosity is notably reduced. Increase the force by which the sample is passed through the needle. Use alternative homogenization methods such as French press or tissue homogenization systems.
18	Inefficient digestion	Protease activity or sample complexity	Adjust the protease concentration to the samples using small scale pilot conditions. The protease activity is maximal at 8-8.5 pH, however, we do not recomment increasing the pH over 7.8 since RNA degradation may occur at higher pH.
18	RNA degradation	RNases in the eluate	We recommend increasing the concentration of RNasin or supplement the buffer with ribonucleoside vanadyl complex (NEB, S1402S)
49	Low peptide yield after MS	Peptides are lost during FASP	RNase treatment can be done after the trypsin treatment. RNA helps to retain the peptides in the filter unit.
49	MS Signal is dominated by tryptic, LysC or ArgC autolysis fragments (e.g. m/z 421.8)	too high concentration of protease	Decrease the amount of the problematic protease. Small scale optimisation experiments will help to define the minimum concentration of protease able to proteolyze efficiently with neglegible miscleavages.
27	Low RNA yield in the second oligo(dT) capture	Low hybridization efficiency	Increase the volume of 1x of hybridization buffer to 10 ml and the amount of beads to 2 ml.
49	Low peptide yield after MS	Peptides are lost during FASP	FASP can be replaced by in solution digestion.
49	Incomplete peptide labeling	Labeling solutions were passing through the StageTips too fast	Adjust the centrifugation speed and duration in order to decrease the flowrate of the labeling solutions through the StageTips.
51	High incidence of infinite and zero ratios	Indentification of low abundant peptides.	Apply a semiquantitative method to determine whether these hits are true or false positives. See reference 15.

REFERENCES

1 Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8**, 479-490 (2007).

2 Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172-177, doi:10.1038/nature12311 (2013).

3 Sampath, P. *et al.* Noncanonical function of glutamyl-prolyl-tRNA synthetase: gene-specific silencing of translation. *Cell* **119**, 195-208, doi:S0092867404009407 [pii]

10.1016/j.cell.2004.09.030 (2004).

4 Ciesla, J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim Pol* **53**, 11-32, doi:20051211 [pii] (2006).
5 Muckenthaler, M. U., Galy, B. & Hentze, M. W. Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. *Annu Rev Nutr* **28**, 197-213, doi:10.1146/annurev.nutr.28.061807.155521 (2008).
6 Chang, C. H. *et al.* Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239-1251, doi:10.1016/j.cell.2013.05.016 (2013).
7 Jarvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. *Cell Commun Signal* **14**, 9, doi:10.1186/s12964-016-0132-3 (2016).
8 Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol Cell* **46**, 674-690 (2012).
9 Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393-1406 (2012).
10 Castello, A. *et al.* System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc* **8**, 491-500 (2013).
11 Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20**, 1122–1130, doi:10.1038/nsmb.2638 (2013).
12 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**, 127-133 (2013).
13 Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* **6**, 10127, doi:10.1038/ncomms10127 (2015).
14 Matia-Gonzalez, A. M., Laing, E. E. & Gerber, A. P. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat Struct Mol Biol* **22**, 1027-1033, doi:10.1038/nsmb.3128 (2015).
15 Sysoev, V. O. *et al.* Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nat Commun* **7**, 12128, doi:10.1038/ncomms12128 (2016).
16 Wessels, H. H. *et al.* The mRNA-bound proteome of the early fly embryo. *Genome Res* **26**, 1000-1009, doi:10.1101/gr.200386.115 (2016).
17 Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S. & Gehring, C. The RNA-binding protein repertoire of Arabidopsis thaliana. *Sci Rep* **6**, 29766, doi:10.1038/srep29766 (2016).
18 Reichel, M. *et al.* In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. *Plant Cell* **28**, 2435-2452, doi:10.1105/tpc.16.00562 (2016).
19 Zhang, Z. *et al.* UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. *Plant Methods* **12**, 42, doi:10.1186/s13007-016-0142-6 (2016).
20 Bunnik, E. M. *et al.* The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. *Genome Biol* **17**, 147, doi:10.1186/s13059-016-1014-0 (2016).
21 Conrad, T. *et al.* Serial interactome capture of the human cell nucleus. *Nat Commun* **7**, 11212, doi:10.1038/ncomms11212 (2016).
22 Gerstberger, S., Hafner, M., Ascano, M. & Tuschl, T. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv Exp Med Biol* **825**, 1-55, doi:10.1007/978-1-4939-1221-6_1 (2014).

23 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845, doi:10.1038/nrg3813 (2014).
24 Choudhury, N. R. *et al.* Trim25 Is an RNA-Specific Activator of Lin28a/TuT4-Mediated Uridylation. *Cell Rep* **9**, 1265-1272, doi:10.1016/j.celrep.2014.10.017 (2014).
25 Popow, J. *et al.* FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and translation. *RNA* **21**, 1873-1884, doi:10.1261/rna.052365.115 (2015).
26 Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcarcel, J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell* **57**, 7-22, doi:10.1016/j.molcel.2014.10.030 (2015).
27 Tejedor, J. R., Papasaikas, P. & Valcarcel, J. Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol Cell* **57**, 23-38, doi:10.1016/j.molcel.2014.10.029 (2015).
28 Fernandez-Chamorro, J. *et al.* Identification of novel non-canonical RNA-binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res* **42**, 5742-5754, doi:10.1093/nar/gku177 (2014).
29 Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell*, doi:10.1016/j.molcel.2016.06.029 (2016).
30 Liao, Y. *et al.* The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Rep*, doi:10.1016/j.celrep.2016.06.084 (2016).
31 Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem Sci* **16**, 323-326 (1991).
32 Schmidt, C., Kramer, K. & Urlaub, H. Investigation of protein-RNA interactions by mass spectrometry--Techniques and applications. *J Proteomics* **75**, 3478-3494, doi:10.1016/j.jprot.2012.04.030 (2012).
33 Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods*, doi:10.1038/nmeth.3092 (2014).
34 Strein, C., Alleaume, A. M., Rothbauer, U., Hentze, M. W. & Castello, A. A versatile assay for RNA-binding proteins in living cells. *RNA* **20**, 721-731 (2014).
35 He, C. *et al.* High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol Cell* **64**, 416-430, doi:10.1016/j.molcel.2016.09.034 (2016).
36 Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416, doi:10.1016/j.cell.2015.03.025 (2015).
37 McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236, doi:10.1038/nature14443 (2015).
38 Rogell, B. *et al.* Specific RNP capture with antisense LNA/DNA mixmers. *RNA* **Accepted** (2017).
39 Liepelt, A. *et al.* Identification of RNA-binding proteins in macrophages by interactome capture. *Mol Cell Proteomics*, doi:10.1074/mcp.M115.056564 (2016).
40 Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **4**, 484-494, doi:10.1038/nprot.2009.21 (2009).

41 Richards, A. L. *et al.* One-hour proteome analysis in yeast. *Nat Protoc* **10**, 701-714, doi:10.1038/nprot.2015.040 (2015).
42 Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169, doi:10.1074/mcp.M400129-MCP200

M400129-MCP200 [pii] (2004).
43 Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
44 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:nbt.1511 [pii]

10.1038/nbt.1511 (2008).
45 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805, doi:10.1021/pr101065j (2011).
46 Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3, doi:10.2202/1544-6115.1027 (2004).
47 Dolken, L. *et al.* High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972, doi:10.1261/rna.1136108 (2008).
48 Radle, B. *et al.* Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *J Vis Exp*, doi:10.3791/50195 (2013).
49 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362, doi:nmeth.1322 [pii]

10.1038/nmeth.1322 (2009).
50 Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906, doi:nprot.2007.261 [pii]

10.1038/nprot.2007.261 (2007).
51 Krijgsveld, J., Gauci, S., Dormeyer, W. & Heck, A. J. In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J Proteome Res* **5**, 1721-1730, doi:10.1021/pr0601180 (2006).
52 Villen, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* **3**, 1630-1638, doi:nprot.2008.150 [pii]

10.1038/nprot.2008.150 (2008).
53 Batth, T. S., Francavilla, C. & Olsen, J. V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* **13**, 6176-6186, doi:10.1021/pr500893m (2014).
54 Hoopmann, M. R. & Moritz, R. L. Current algorithmic solutions for peptide-based proteomics data generation and identification. *Curr Opin Biotechnol* **24**, 31-38, doi:10.1016/j.copbio.2012.10.013 (2013).
55 Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371, doi:nprot.2009.2 [pii]

10.1038/nprot.2009.2 (2009).

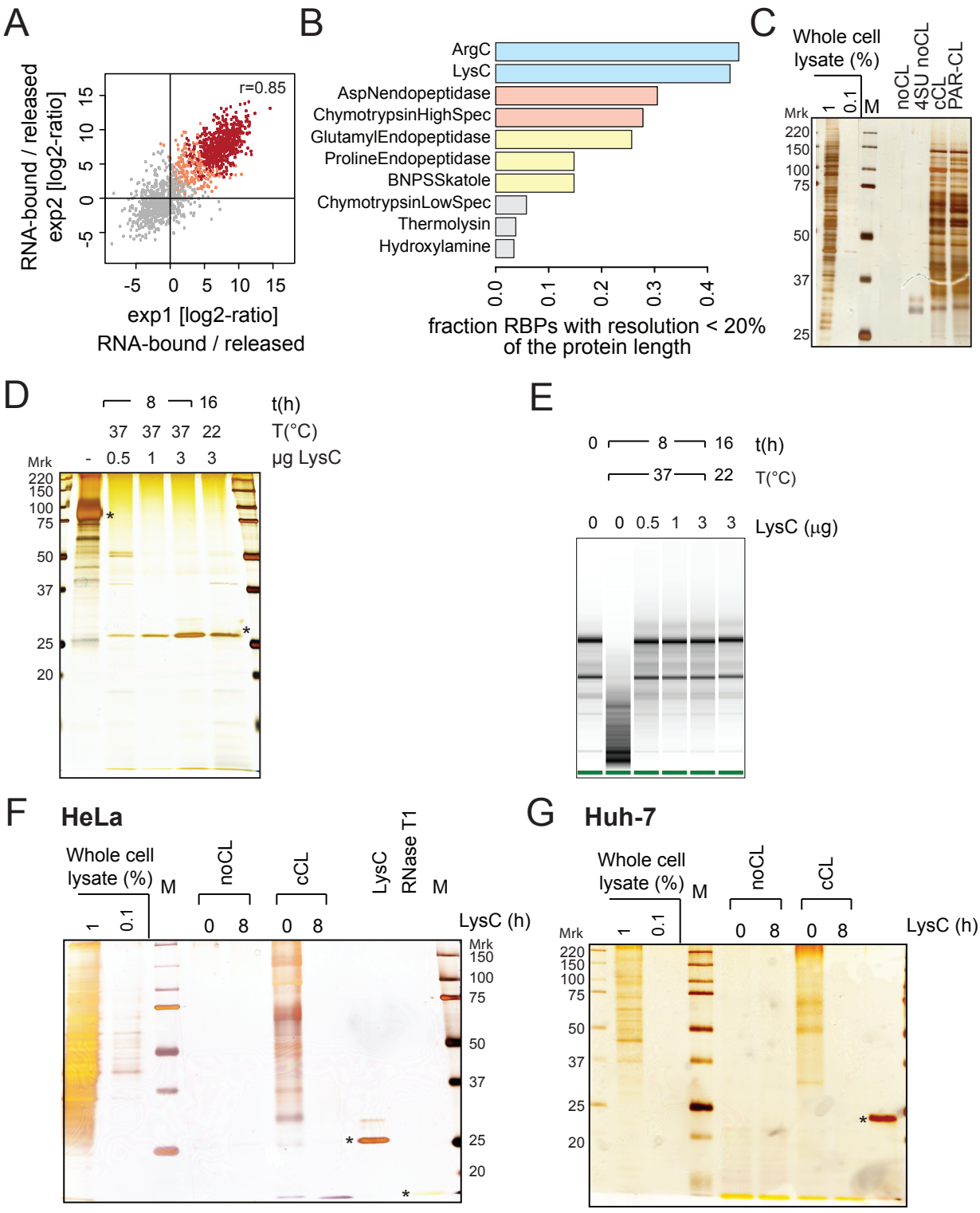


Figure 2 (Castello et al)

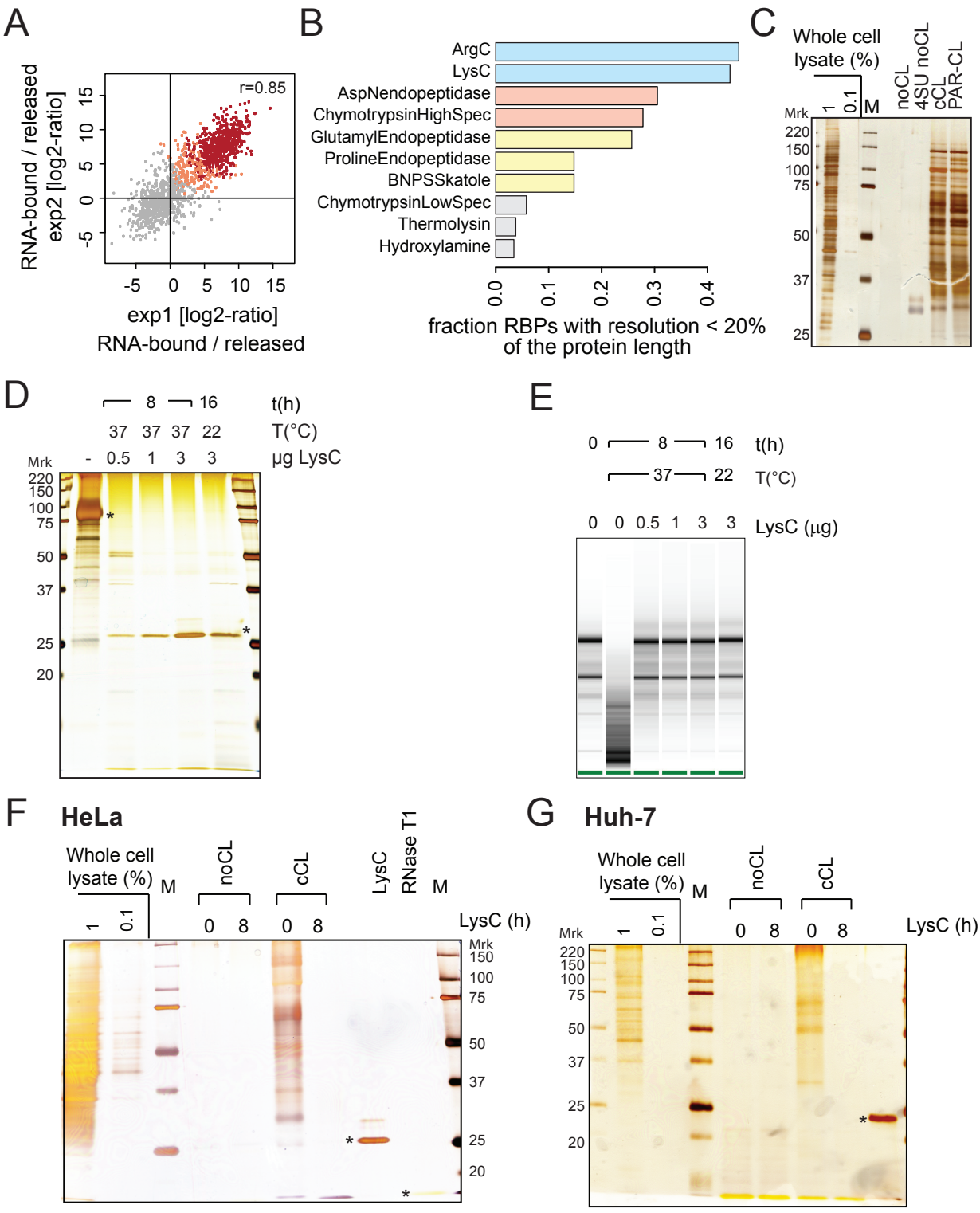


Figure 2 (Castello et al)

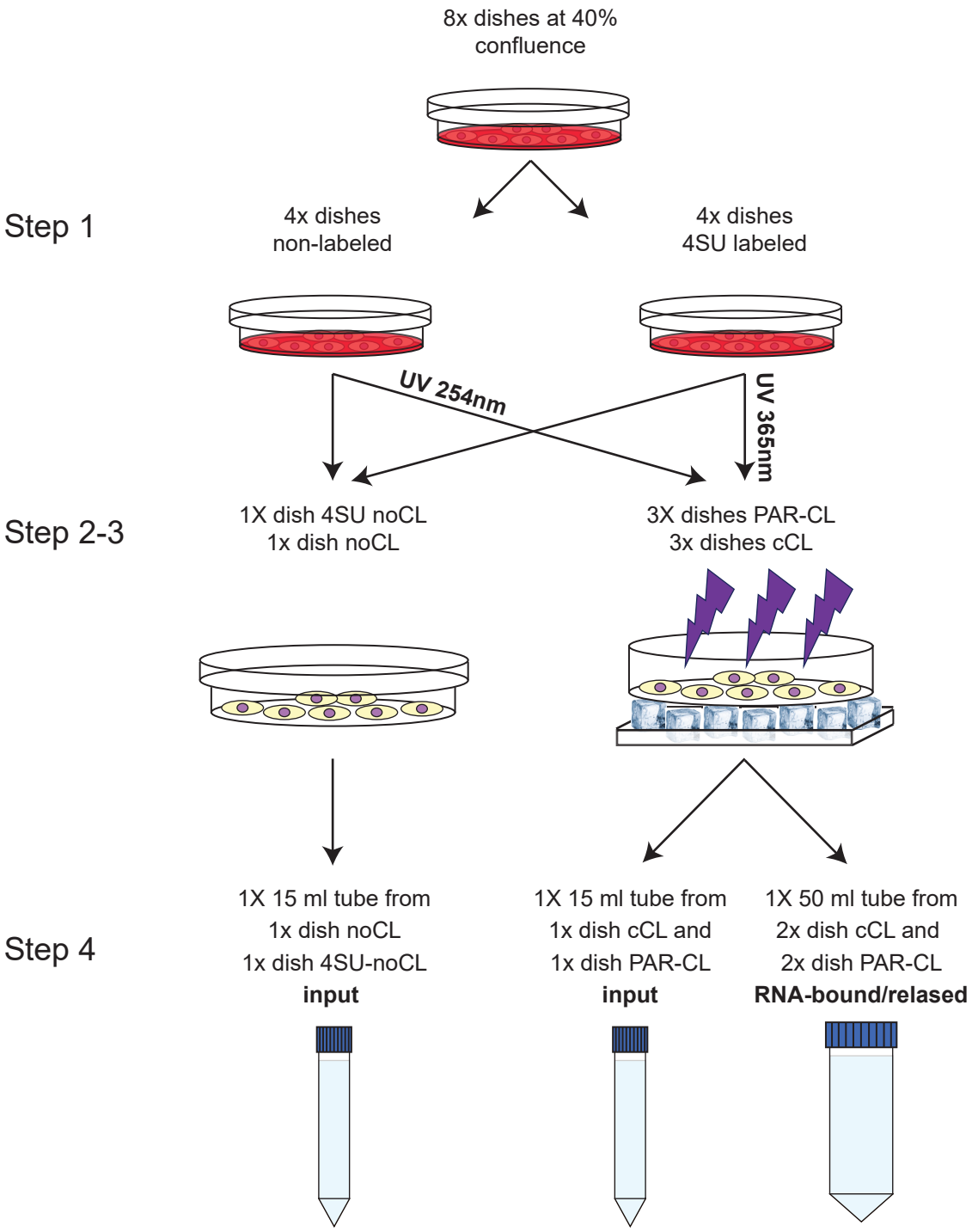
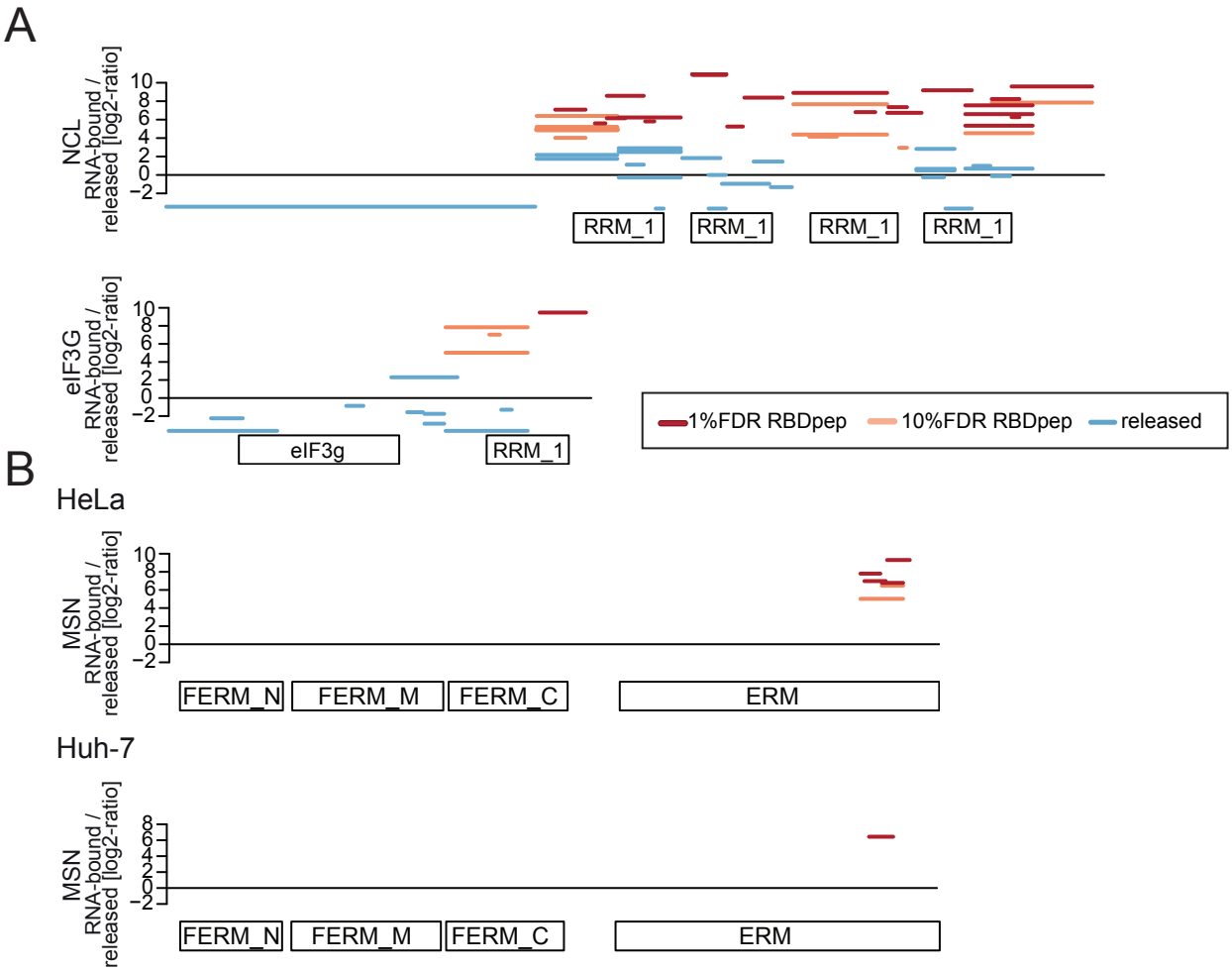


Figure 3 (Castello et al)



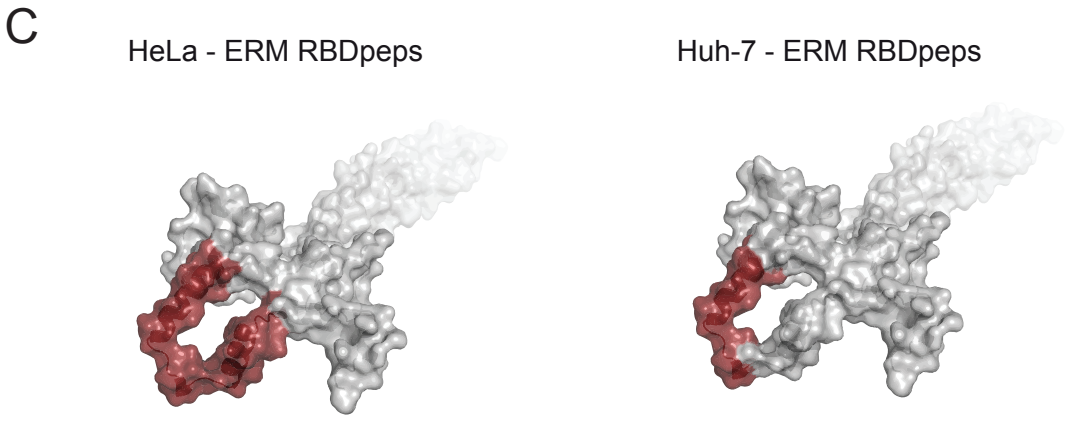


Figure 4 (Castello et al)